

Which health and biomedical topics generate the most Facebook interest and the strongest citation relationships?*

Ehsan Mohammadi^{1*}, Karl Gregory², Mike Thelwall³, Nilofar Barahmand⁴

^{1*}College of Information and Communications, University of South Carolina, Columbia, SC, USA

²Department of Statistics, University of South Carolina, Columbia, SC, USA

³Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wolverhampton, UK

⁴Scientometrics Division, Shiraz University of Medical Sciences, Shiraz, Iran

Highlights

- Four biomedical areas, including General and Internal Medicine and Complementary and Alternative Medicine receive more Facebook attention.
- No biomedical areas receive little Facebook attention.
- The relationship between Facebook mentions and citations varies by field and public interest.
- Facebook mentions predict future citations less well for fields with the variable public interest.

Abstract

Although more than a million academic papers have been posted on Facebook, there is little detailed research about which fields or cross-field issues are involved and whether there are field or public interest relationships between Facebook mentions and future citations. In response, we identified health and biomedical scientific papers mentioned on Facebook and assigned subjects to them using the MeSH and Science Metrix journal classification schema. Multistage adaptive LASSO and unpenalized least-squares regressions were used to model Facebook mentions by fields and MeSH terms. The fields Science and Technology, General and Internal Medicine, Complementary and Alternative Medicine, and Sport Sciences produced higher Facebook mention counts than average. However, no MeSH cross-field issue differences were found in the rate of attracting Facebook mentions. The relationship between Facebook mentions and citations varies between both fields and MeSH cross-field issues. General and Internal Medicine, Cardiovascular System and Hematology and Developmental Biology have strongest correlations between Facebook mentions and citations, probably due to high citation rates and high Facebook visibility in these areas.

**This is a preprint of an article to be published in the Information Processing and Management© copyright 2020 Elsevier.

1. Introduction

Assessing the societal impact of scientific outputs is important for universities, science funders, and policymakers to understand the influence of science beyond academia and to demonstrate the value of funded projects to taxpayers and donors (Abramo, 2018). For instance, the National Institutes of Health (NIH), as the largest public funder of health and medical sciences, is interested in both scientific impact and the impact of biomedical publications on society (National Institutes of Health, 2019). Although many organizations have attempted to capture the societal impacts of health publications, this is not straightforward because of the many ways in which impact can occur (Bornmann, 2013), and so multiple approaches are needed. For example, many studies have tried to use social media to reveal aspects of the impact of science outside academia (Bornmann, 2013; Mohammadi, Thelwall, Kwasny, & Holmes, 2018; Holmberg, Bowman, Bowman, Didegah, & Kortelainen, 2019). Facebook may be able to provide evidence of public interest in research because it claims to have 2.4 billion active members (Facebook, 2019), which is half of the world's internet users (Internet World Stats, 2019). These can interact to share, comment and endorse research and other ideas. As a result, Facebook has a much wider potential reach than academic social network sites like ResearchGate, Academia.edu, and Mendeley, and is therefore a better choice for societal impact evidence. Although Facebook mentions of academic papers might reflect non-academic impacts, there is insufficient evidence to draw this conclusion because they might also be spam or publicity from the publishing journals or authors.

Health-related research is probably more interesting for the public than other fields because it is more directly relevant to their lives. There is extensive media coverage of diseases like cancer (Lewison & Sullivan, 2008) and mental disorders (Lewison, Roe, Wentworth, & Szmukler, 2012), for example. As a result, Facebook could also be used to disseminate health-related information to the public. A content analysis of a small sample of psychology papers shared on Facebook confirmed that the public could engage with scholarly information (Na & Ye, 2017). Particularly, a case study about Zika virus revealed that Facebook might be a better platform than Twitter to share scientific findings to relevant populations (Barata, Shores, & Alperin, 2018). Nevertheless, the role of Facebook in sharing medical science articles is not well understood. To fill the gap, this paper covers a broad sample of scientific papers shared on Facebook across different biomedical subjects from different perspectives.

When analyzing academic impact, it is important to split articles by their academic fields because different fields can have different types of impact. This is usually achieved by exploiting a journal classification system rather than attempting to classify individual papers (Archambault, Beauchesne, & Caruso, 2011), although the latter is probably more accurate (Shu et al., 2019). With the growth of technology, it has become easier to obtain classifications for individual articles, however (Fenner, 2013). The United States National Library of Medicine (NLM) maintains the

Medical Subject Headings (MeSH) classification schema for the main topics of individual biomedical articles covered by PubMed, using robust semantic hierarchical relationships between concepts (Liu & Wacholder, 2017). MeSH is therefore a useful classification source for investigating the impact of biomedical research for individual articles. Despite this, MeSH has rarely been used for evaluating research publications (Leydesdorff & Opthof, 2013). MeSH has the additional advantage that it is multifaceted and not subject-based. Its classifications can therefore be used to delineate sets of articles in a different way, giving the potential in the current paper to reveal areas of apparent public interest on Facebook that do not correspond to academic fields. The primary objective of the current paper is therefore to exploit both journal-based and MeSH classifications in order to identify fields and cross-field issues of public interest about biomedical research on Facebook.

Previous studies have found low correlations between Facebook mentions of academic papers and citation counts overall (Haustein, Costas, & Larivière, 2015; Thelwall, Haustein, Larivière, & Sugimoto, 2013), but the extent to which the relationship between citations and Facebook mentions varies between subjects is not known. Thus, another objective of this study is to investigate this issue systematically for individual subjects within the biomedical domain.

2. Research questions

The primary goal was to assess whether the amount of attention on Facebook varied between fields, as reflected by a standard journal classification scheme, and between cross-field issues, as reflected by MeSH. The secondary goal was to investigate topic and issue variations in the relationship between citations and Facebook mentions. An additional methodological goal was to investigate whether combining journal-level and article-level classifications can give finer-grained information for health-related research.

RQ1: Do any medical fields and cross-field issues (as reflected by relevant MeSH terms) generate above or below average interest on Facebook, as indicated by mentions?

RQ2: Does the relationship between Facebook mentions and citation counts vary substantially between medical fields and cross-field issues (as reflected by relevant MeSH terms)?

3. Literature review

3.1. Article and Journal level classifications

MeSH is a controlled vocabulary of biomedical concepts maintained by the US National Library of Medicine (NLM). MeSH is a semantic ontology in a hierarchical structure (Kastrati, Imran, & Yayilgan, 2019). Human indexers at NLM define and maintain MeSH conceptual relations using

rules in 16 categories (Anatomy, Organisms, Diseases, Chemicals and Drugs, Analytical, Diagnostic and Therapeutic Techniques and Equipment, Psychiatry and Psychology, Phenomena and Processes, Disciplines and Occupations, Anthropology, Education, Sociology and Social Phenomena, Technology, Industry, Agriculture, Humanities, Information Science, Named Groups, Health Care, Publication Characteristics, Geographicals), each of which has a basic tree structure with up to thirteen hierarchical levels (see https://www.nlm.nih.gov/mesh/intro_trees.html). Each article in PubMed is typically assigned several MeSH terms. These can be used in the PubMed search interface for MeSH-specific queries.

Although designed for indexing and information retrieval, MeSH terms have been used in scientometric research, such as to map the landscapes of health and medical sciences fields (Leydesdorff & Opthof, 2013; Leydesdorff, Comins, Sorensen, Bornmann, & Hellsten, 2016). Due to the lack of links between the Medline search interface for PubMed and the major international citation indexes, few studies have connected MeSH data with citation counts to investigate the citation impact of biomedical papers using MeSH terms (Leydesdorff & Opthof, 2013). The MH2WoS application (<https://www.leydesdorff.net/software/scopus.htm>) has been developed to partly overcome this barrier. A study using this tool revealed that journal categories and MeSH terms cover different attributes of academic papers about Alzheimer's disease (Leydesdorff et al., 2016). Another study found that Neoplasms, Tuberculosis, and Disease were the most common MeSH terms in biomedical papers cited by Wikipedia (Dehdarirad, Didegah, & Sotudeh, 2018).

There are different methods for assigning subjects to journals in academic databases, such as the schemas developed by the Web of Science, Scopus, Excellence in Research for Australia, and the National Science Foundation. In these systems, journals are classified rather than papers because it more practical. This is an imperfect solution because of interdisciplinary journals and because some documents are published in journals classified into unrelated subject categories (Gómez, Bordons, Fernández, & Méndez, 1996). Additionally, emerging topics may not fit into existing classifications. The Science-Metrix team has compared different journal classification systems and proposed an inclusive journal-level schema (Archambault et al., 2011) which includes six broad domains, 22 fields and 176 subfields. We adopted Science-Metrix scheme because of this and because each journal is categorized into a single category. This complements the MeSH scheme so that articles misclassified by Science-Metrix due to being published in interdisciplinary journals may be better classified by MeSH.

3.2. Academic publications on Facebook

Facebook is the social media platform most used in academia (Enkhbayar, Haustein, Barata, & Alperin, 2019; Bowman, 2015). In addition, 80% of the medical journals in one study had a Facebook presence (Kamel Boulos & Anderson, 2012), as did 80% of the biomedical journals in another (Zedda & Barbaro, 2015). Since a cross-disciplinary study revealed that only 7% to 14%

journals had Facebook accounts, journals in art and humanities had more coverage (Zheng et al., 2019), this suggests that medical and biomedical journals are particularly likely to be on Facebook.

A minority of scientific articles seem to be mentioned on Facebook, with disciplinary and time variations. The Facebook coverage of scientific papers varies based on data aggregators, however (Zahedi & Costas, 2018). A cross discipline analysis found 2.5% of Web of Science scholarly articles with a DOI published mid-2011 to 2013 to be mentioned on Facebook by 2013, with most mentions occurring for the medical and life sciences and natural sciences (Costas, Zahedi, & Wouters, 2015). Similarly, 3% of Latin American journal articles are posted on Facebook (Alperin, 2015). Some psychology fields, such as applied psychology, are mentioned more on Facebook than others, including mathematical psychology (Na & Ye, 2017). Another multidisciplinary study confirmed that health and biomedical articles were more likely to be mentioned on Facebook (7.5%) than the social and earth sciences (Haustein et al., 2015). For *Nature* articles 2010 to 2015, Facebook coverage increased over time, with most mentions being related to human health (Xia et al., 2016). The greater prevalence of health and biomedical articles on Facebook could be due to the greater presence of journals in these areas, assuming that they publicize their articles on the site.

Facebook mentions have also been proposed as a societal impact indicator for academic research on the basis that mentions may reflect public interest in, or use of, published research. The first approach to assess a proposed new research impact indicator is to measure its correlation with citation counts on the basis that positive correlations indicate non-random data and even societal impacts should be related to academic impacts (Sud & Thelwall, 2014). Several studies have found very weak correlations between Facebook mentions and citation counts, however, but stronger correlations with Twitter (Haustein et al., 2015; Thelwall et al., 2013). No prior study has reported the relationship between citations and Facebook mentions for individual fields, however, so it is not known if there are any fields with strong relationships. The low correlations were partly due to the low numbers of Facebook mentions, in comparison to tweets. An analysis of publications submitted to the UK Research Excellence Framework 2014 showed that altmetrics including Facebook mentions had statistically insignificant correlations with REF reviewer judgements (HEFCE, 2015). Nevertheless, one study of psychology papers discovered that most users who shared them on Facebook were non-academics (Na & Ye, 2017), giving hope that Facebook mentions might be indicators of public interest.

3.3. Public interest in biomedical information

It is not straightforward to understand the role of the media in shaping public opinion, although informing citizens can be a goal of some media agencies (Rees & Bath, 2000). This is particularly true for health-related news (Collins, Abelson, Pyman, & Lavis, 2006). News platforms can play a mediating role in disseminating and translating scientific findings for the public, policymakers and medical professionals (Noyons, 2019). Few academic papers are selected for press coverage,

with the focus being presumably on engaging topics. Nevertheless, there seems to be an increasing use of academic research as the basis for news stories (Kousha & Thelwall, 2019).

An analysis of European news stories related to research indicated that diabetes was extensively discussed on social media (Pallari et al., 2018). Five UK newspapers gave medical topics more attention than other science areas (Weitkamp, 2003). Also in the UK, the research topics most discussed in media were related to diet, oncology, gynecology, and pharmacy (Lewison, 2002). An analysis of two British newspapers found that they had cited 7% of Lancet and British Medical Journal articles, with topics mainly related to female health and cancer (Bartlett, 2002). Another study reported that neuroscience and genetics were the main biomedical research subjects in the *New Scientist* magazine (Grant Lewison & Turnbull, 2010). Recently, new methods were used to analyze large-scale press data and reported that medical journals were the most cited by eight British newspapers (Kousha & Thelwall, 2019).

3.4. Factors influencing social media metrics and citations

Factors that may influence journal article citation counts may also influence media mentions. These include multiple authors (Didegah, 2014), university rank (Leimu & Koricheva, 2005), international collaboration (Persson, 2010) and journal prestige (Stremersch, Verniers, & Verhoef, 2007). A study comparing the differences between the factors influencing citations and social media metrics for academic papers found that collaboration and the number of references were the main common factors, although the pattern for Facebook was minimal in biomedical and health articles (Haustein et al., 2015). Another study found that inter-institutional and national collaboration associated with decreases in social media mentions whilst international collaboration associated with more Twitter mentions and Mendeley readers, but not more Facebook mentions (Didegah, Bowman, & Holmberg, 2016). This study also found that social sciences articles had received more attention on Twitter and Facebook than papers from the physical and natural sciences. The journal impact factor is also related to the number of Facebook mentions for papers published by Finnish scholars (Didegah, Bowman, & Holmberg, 2018).

One of the advantages of social media mentions is that they appear before citations (Mohammadi et al., 2018) and some studies have focused on understanding the relationship between various social media metrics and future citations. There is some evidence to show that Twitter mentions (Eysenbach, 2011), Mendeley readers, (Thelwall & Nevill, 2018) and blog posts (Shema, Bar-Ilan, & Thelwall, 2014) of academic papers can predict their future citations in a wide range of academic fields. Facebook likes can help predict future citations for psychology articles but not for other areas, such as business and the life sciences (Ringelhan, Wollersheim, & Welp, 2015).

4. Methods

The research design was to gather a large set of biomedical papers mentioned on Facebook and to fit linear regression models for the log-normalized citation count and Facebook mention variables, incorporating MeSH terms, and field as independent variables.

4.1. Data

Facebook mentions of academic papers were extracted from an Altmetric.com database snapshot, which includes Facebook mentions collected by Altmetric.com until July 2018. We searched this dataset for records with a DOI and at least one Facebook mention. Out of 19,404,205 records, there were 1,316,210 matches. This is zero truncated data in the sense that papers without Facebook mentions were excluded. Altmetric.com papers without Facebook mentions were ignored because these papers would have had other web mentions to be tracked by Altmetric.com and so would be unrepresentative of papers ignored by Facebook. Thus, the analysis here is about the *intensity* (number) of Facebook mentions for papers mentioned on Facebook rather than whether papers are mentioned on Facebook.

Papers written before 2014 were excluded to focus on years where Altmetric.com was actively collecting data for the complete year so that the data would be comparable (Altmetric data for older papers would be retrospective Facebook mentions rather than publication-time mentions). The fields of each paper were obtained by matching their journal information with the Science-Metrix classification scheme (<http://science-metrix.com/?q=en/classification>) which assigns each journal to a single subject category from a list of 176 (Archambault et al., 2011). We used International Standard Serial Number (ISSN) and electronic ISSN to match our data with Science-Metrix classification. Citation counts for these papers were obtained from the Dimensions citation index, which is a reliable resource with coverage comparable to Scopus and probably more extensive than the Web of Science (Thelwall, 2018).

We searched the Almetric.com papers with Facebook mentions in PubMed using PMIDs to add MeSH data, finding 134,052 matches. Detailed information about the distribution of the records across different health biomedical fields is shown in Appendix 1. Each paper could be expected to have 10 to 12 MeSH terms for the key topics at different hierarchical levels (Chapman, 2009). Although terms occur at different levels in the MeSH hierarchy, we used only second-level terms (n=118; e.g., Cardiovascular System [A07]) as a compromise between very general and quite specific descriptors. MeSH terms that align with fields provide article-level field classifications that would be most powerful for papers in multidisciplinary journals (e.g., Science). The remaining MeSH terms provide cross-field descriptors that could be thought of as indicating issues that are common to multiple fields. MeSH does not systematically distinguish between the two types and so human judgement is needed to assess how individual MeSH terms in the results should be interpreted.

The final data set consisted of 134,052 records of papers published 2013-2017, with at least one Facebook mention recorded by Altmetric.com by July 2018, with citation counts from Dimensions, in a journal with a Science-Metrix classification (one per paper), and with a PubMed ID and multiple article-level second-level MeSH terms from PubMed. The dataset includes multiple document types, such as standard articles, reviews, letters and editorials, but standard articles and reviews probably dominate (we did not have document type data) if other document types are less likely to be mentioned on Facebook. The inclusion of multiple document types is a limitation of the method. Appendix 2 shows number of articles in the final data set based on the second level MeSH terms. All terms second level MeSH terms are in the datasets except V01-V04, which are publication-related characteristics.

Some summary statistics from the Facebook mention and citation data are given in Table 1, which shows several percentiles of the distributions as well as the maximum, the mean, and the standard deviation. At least half of the papers were mentioned on Facebook only once, and at least half of the papers were cited 10 or fewer times. Only 0.5% of the papers were mentioned on Facebook more than 41 times, and only 0.5% of the papers were cited more than 418 times. We see that the distributions of the Facebook mentions and the numbers of citations are highly right-skewed. For this reason, we focus on log transformations of these variables, which reduces the influence of high outliers on our analyses.

Table1. Several percentiles of the distributions of Facebook mentions and citations from the dataset.

	5%	10%	25%	50%	75%	90%	99%	99.5%	Max	Mean	Sd
Facebook Mentions	1	1	1	1	2	5	26	41	2341	2.90	14.26
Citations	0	1	4	10	25	56	281	418	7140	26.95	79.86

4.2. Models

We addressed the research questions by fitting regression models with some or all of the following variables.

- C is the number of times the paper has been cited and $\tilde{C} = \log(C + 1)$. Log transformations were used because citation data is highly skewed.
- F is the number of times the paper has been mentioned in a Facebook post and $\tilde{F} = \log(F + 1)$. Although all papers have at least one Facebook mention, by design, one is added to match up with the citation counts and to mirror standard practice.
- $\text{MeSH}_1, \dots, \text{MeSH}_J$ are binary indicators for the attribution of each of $J = 141$ different MeSH terms to the paper, such that $\text{MeSH}_j = 1$ if MeSH term j was assigned to the paper, otherwise 0.

- $\text{Field}_1, \dots, \text{Field}_K$ are $K = 92$ different Science-Metrix field membership indicators, such that $\text{Field}_j = 1$ if the paper is in field j , otherwise 0. The field **Fluids and Plasma** was chosen as the baseline in initial model fitting solely due to its position in the data set as the 93rd field; the estimated models based on this choice contained very large numbers of insignificant field effects as well as a small number of highly significant field effects. This indicated that most of the fields were not significantly different from **Fluids and Plasma**. As a result, we kept this choice of the baseline and have regarded it in our interpretations as representative of a mean level over all fields with insignificant deviations from it, and we have interpreted the highly significant effects as representing deviations from this mean level.
- $\text{Year}_1, \dots, \text{Year}_T$ are binary indicators of publication year, such that $\text{Year}_j = 1$ if the paper was published in the j th year otherwise 0, for $j = 1, \dots, T = 4$. The baseline year is 2017, so articles published in 2017 have zeros on all four year indicators.

Some of the regression models include penalized variables. These are only included in the final model if their effect sizes are substantial. The multistage adaptive LASSO variable selection technique (Tibshirani, 1996; Bühlmann & Meier, 2008) was used to choose which coefficients among the penalized groups are nonzero, with the remaining terms being removed from the model. A variable selection procedure is necessary because the very large sample size, $n = 133470$, would otherwise probably give many highly statistically significant p -values, causing many predictors with small effects to be retained in the model, despite being of little practical value. This is an iterative procedure and 5 reweighting steps were used.

After selecting variables with the multistage adaptive LASSO, unpenalized least-squares regression was fitted using only the selected covariates in order to give relatively straightforward results and confidence intervals. Confidence intervals were adjusted for multiple testing by controlling the familywise error rate at 0.01 with a Bonferroni correction. For example, this gives the significance level $0.01/237 = 0.000,042$ for the first model, which contains 237 parameters about which we wish to make inferences, and significance level $0.01/475 = 0.000,021$ for the second model, which contains 475 parameters about which we wish to make inferences.

4.2.1 Facebook mentions by fields and MeSH terms

The following model to investigate which MeSH terms and fields attract different numbers of Facebook mentions addresses the first research question. Only the MeSH terms are penalized to ensure that no field differences are ignored when investigating MeSH differences.

$$\log(F + 1) = \mu + \overbrace{\sum_{j=1}^J \alpha_j \text{MeSH}_j}^{\text{penalize}} + \sum_{j=1}^K \gamma_j \text{Field}_j + \sum_{j=1}^T \eta_j \text{Year}_j$$

4.2.2 Citation counts by Facebook mentions, fields and MeSH terms

The following linear regression model was constructed for the log-transformed number of citations papers receive as a function of their log-transformed Facebook mentions, fields, MeSH terms, publication years (dummy variables), and Facebook mention interaction terms. This addresses the second research question. Citations are regressed against Facebook mentions because paper sharing on Facebook is likely to occur before citations appear, so Facebook mentions are more likely to influence citations than the other way around for typical papers, although both causal directions are likely to occur as well as indirect relationships. For example, researchers might post about a paper and cite it, with the citation appearing in the Web of Science years later when their paper is published. Similarly, papers might sometimes be cited and posted on Facebook because they are useful, with no direct causal connection between the two activities. Finally, an interesting paper might be cited and posted about by two different people unaware of each other.

$$\tilde{C} = \mu + \beta \tilde{F} + \overbrace{\sum_{j=1}^J \alpha_j \text{MeSH}_j}^{\text{penalize}} + \sum_{j=1}^K \gamma_j \text{Field}_j + \sum_{j=1}^T \eta_j \text{Year}_j +$$

$$\overbrace{\sum_{j=1}^J \beta_j (\tilde{F} \times \text{MeSH}_j) + \sum_{j=1}^K \zeta_j (\tilde{F} \times \text{Field}_j) + \sum_{j=1}^T \theta_j (\tilde{F} \times \text{Year}_j)}^{\text{penalize}}$$

5. Results

The model fitting results predominantly apply to the health and biomedical fields that are mainly represented in the data (Appendix 1). Fields with few articles, such as Human Factors (n=91), are effectively ignored by the models because of the dual conservative variable selection mechanisms: Bonferroni (all variables) and LASSO (MeSH and interaction terms). Overall, the greater the field sample size, the smaller the effect sizes needed to be statistically significant.

5.1. Facebook mentions by fields and MeSH terms

For Model 1, the multistage adaptive LASSO discards all MeSH terms, implying that second-level article-level MeSH terms are not substantial predictors of the number of Facebook mentions (for papers receiving at least one Facebook mention), after accounting for journal-level Science-Metrix fields. Model 1 suggests that only four fields have substantially more logged Facebook mentions than the mean of the other fields (Figure 1).

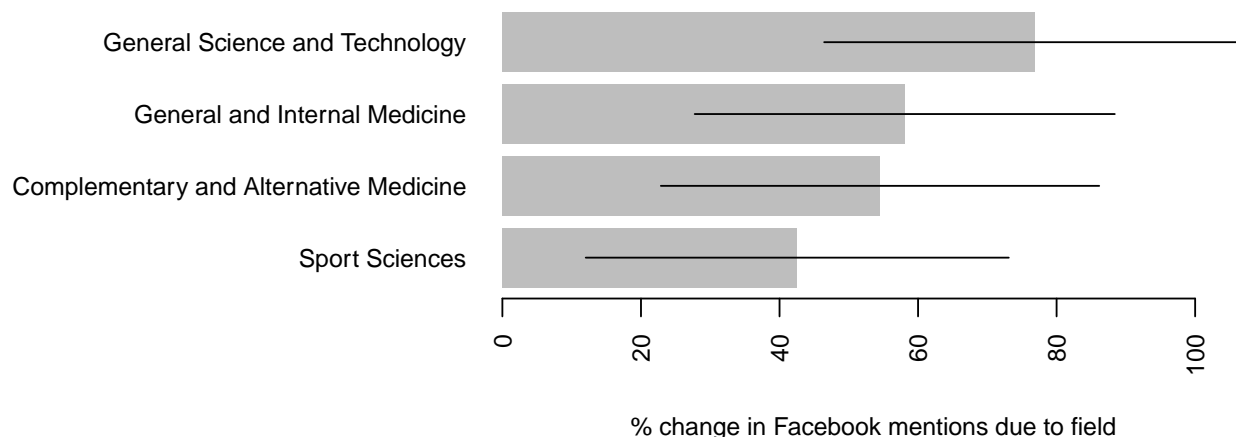


Figure 1: Model 1 estimated % change in logged Facebook mentions due to field membership (terms shown when 99% Bonferroni-adjusted ($n=237$) confidence interval excludes 0).

The results revealed that four fields, *Sport Sciences*, *Complementary and Alternative Medicine*, *General and Internal Medicine*, and *General Science and Technology* all have positive relationships with the number of Facebook mentions.

5.2. Citation counts by Facebook mentions, fields and MeSH terms

The Model 2 coefficient $\beta = 0.31$, 99% CI: (0.23, 0.39) represents the effect of logged Facebook mentions on logged times cited. Thus, a 1% increase in logged Facebook mentions, all else held constant, associates with a mean increase of 0.31 in logged citations.

Mean citation counts are known to vary substantially by field and so it is unsurprising that Model 2 found many field differences (marginal effects). These included 20 field differences that were statistically significant, having Bonferroni-adjusted p-values less than 0.01 (given in Tables 4 to 7 of the Appendix 3), for influencing citation rates. Since these occur despite accounting for Facebook mentions, this confirms that there is not a simple cross-field relationship between citation counts and Facebook mentions (i.e., knowing the field of a paper rather than just its Facebook mention count would help when estimating its citation count). **Nanoscience** ($n=691$ papers) has the largest positive coefficient (0.85: relatively many logged citations for the number of logged Facebook mentions) and **History of Science, Technology and Medicine** ($n=62$) has the largest negative coefficient (-1.1), although based on a small sample size.

Seventeen MeSH terms associate with increased citations, irrespective of Facebook mentions and field, and two associate with the opposite (Figure 2). Thus, although no MeSH terms seem to be associated with differing numbers of Facebook mentions after accounting for field differences (Model 1), for some there is a stronger relationship between Facebook mentions and citation counts after accounting for field differences. For example, papers categorized with the MeSH term *Microbiological Phenomena* can expect to have over 30% more logged citations than a comparable paper without this MeSH term. The *Microbiological Phenomena* MeSH term is related to the Science-Metrix field **Microbiology**.

Thus, the power of the *Microbiological Phenomena* MeSH term is presumably in classifying microbiology papers not published in Science-Metrix **Microbiology** journals because they are in general (e.g., Science, Nature) or multidisciplinary (e.g., Lancet, Bmj) journals in other Science-Metrix categories.

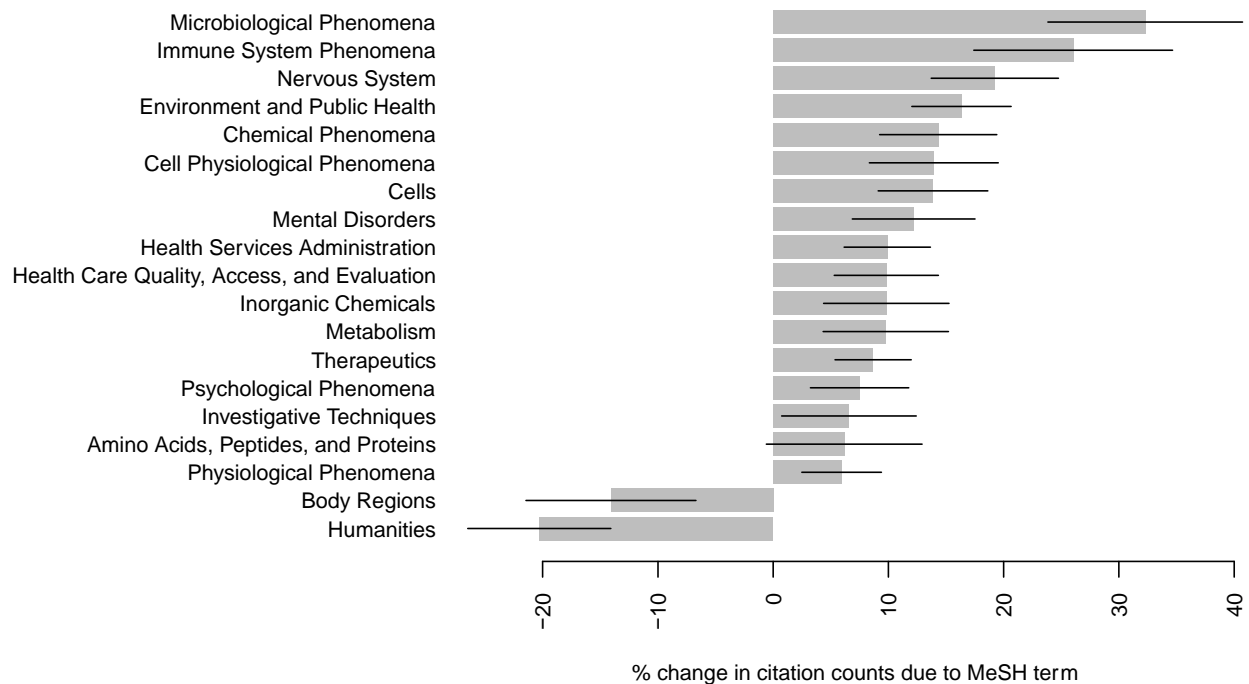


Figure 2: Model 2 MeSH term marginal effects. Estimated % change in citations due to MeSH terms (irrespective of Facebook mentions). Error bars represent 99% Bonferroni-adjusted (n=475) confidence intervals.

There are some significant interaction terms in Model 2 for Facebook mentions and fields (Figure 3). In **General and Internal Medicine** the relationship between Facebook mentions and citation counts is stronger (i.e., each additional Facebook mention of papers in this field is associated with a greater increase in citations than the mean increase for the papers in the dataset), while in **Biochemistry and Molecular Biology** it is weaker.

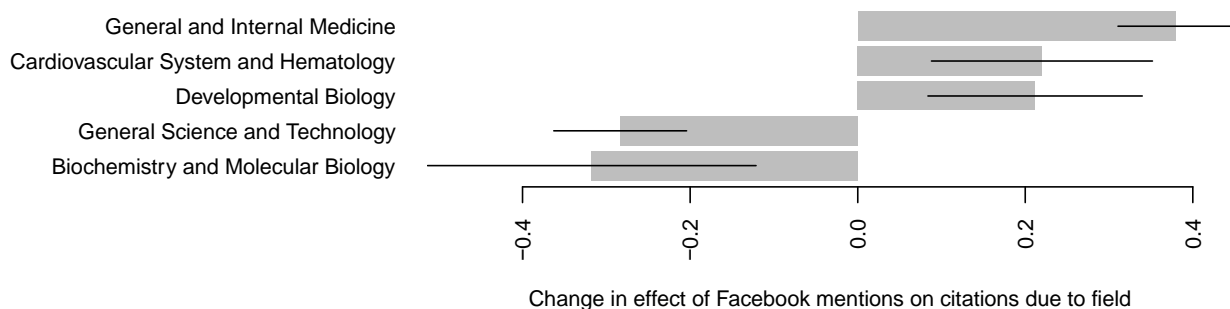


Figure 3: Model 2 estimated change in effect of Facebook mentions on citations due to field membership (interaction effects of field terms and logged Facebook mentions). Error bars represent 99% confidence intervals (terms shown when 99% Bonferroni-adjusted (n=475) confidence interval excludes 0).

There are some significant interaction terms in Model 1 for Facebook mentions and MeSH terms (Figure 4). For example, the MeSH term *Biological and Dental Materials* associates with a stronger positive relationship between Facebook mentions and times cited, whereas the relationship tends to be weaker under the MeSH term *Geographic Locations*.

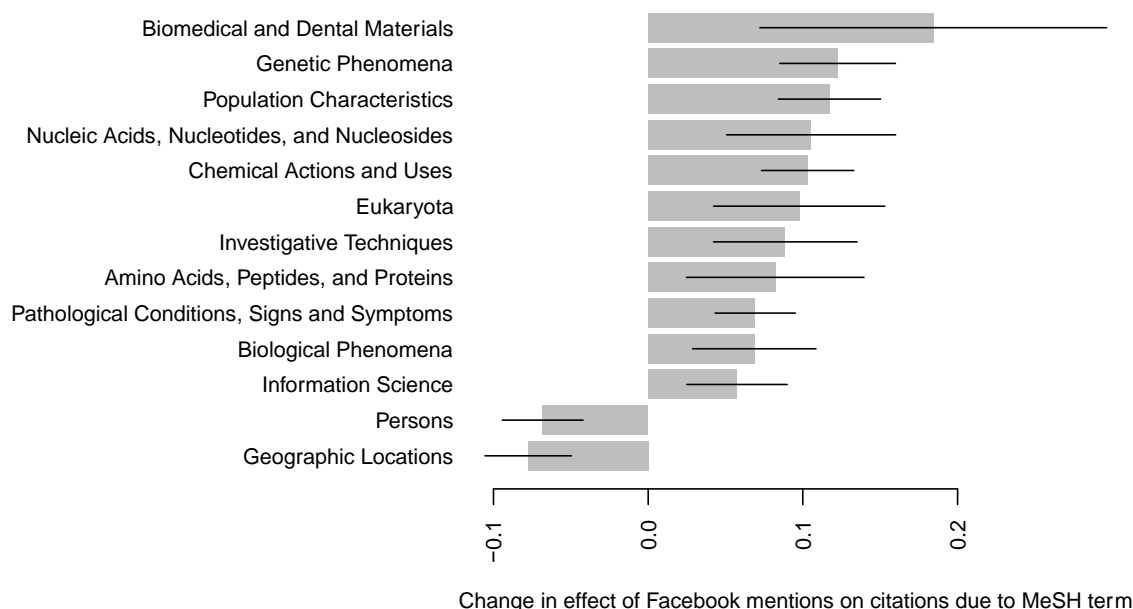


Figure 4: Model 2 Estimated change in effect of Facebook mentions on citations due to MeSH terms (interaction effects of MeSH terms and logged Facebook mentions). Error bars represent 99% confidence intervals (terms shown when 99% Bonferroni-adjusted ($n=475$) confidence interval excludes 0).

5. Discussion

This study combines different data sources and statistical models to get insights into the factors that underlie sharing health and biomedical papers on Facebook, and citation counts for shared biomedical papers. The analysis includes journal-level Science Metrix fields, paper-level MeSH subject headings and citation counts.

RQ1 (factors influencing Facebook mentions): Only four fields and no cross-field issues (from MeSH terms) were found by Model 1 to produce higher Facebook mention counts than the mean. The **General Science and Technology** field includes many large high profile multidisciplinary scientific journals, including *Nature*, *Science* and *PLOS One*. These are known to have high social media visibility (Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014; Vogl, Scherndl, & Kühberger, 2018) and on Facebook in particular (Xia et al., 2016). The high Facebook visibility of **General and Internal Medicine** journal papers is in line with a previous study that found journals in this field to have higher coverage in Facebook than other clinical medicine journals (Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014). This diverse area covers a wide range of illness related to multiple organs and might be interesting to the public because it discusses general issues rather than individual diseases. In contrast, the high profile of **Complementary and Alternative Medicine** journal papers does not seem to have been noticed before. This may be due

to the public using social media to share information about alternative treatments (Sharma, Holmes, & Sarkar, 2016), given that doctors may not recommend or discuss them. **Sport Sciences** journal papers are also highly shared on Facebook, perhaps reflecting public interest in sport. A previous study also found Facebook to be a useful platform for scholars to inform professionals (Williams, 2011). Thus, overall, public interest seems to be a plausible explanation for the most frequently mentioned four fields. This is reasonable because most Facebook users that share academic papers are not academics (Mohammadi, Barahmand, & Thelwall, 2019). This study contributes to knowledge by identifying Facebook-friendly academic fields (e.g. **Complementary and Alternative Medicine**) for the first time.

The lack of cross-field issues is surprising given that previous research into press coverage of medical research has found recurrent topics of interest, such as cancers and psychological illnesses, none of which were found to associate with more Facebook mentions. This is not due to data sparseness, since **Oncology & Carcinogenesis** is one of the largest topics (n=5179 papers analyzed, but not significant in Model 1), as is **Psychiatry** (n=6463, not significant) and **Clinical Psychology** (n=1128, not significant) is not small. This might be due to Science-Metrix fields mixing generally relevant (e.g., cancer cures) and specialist interest (e.g., cancer spread modelling) so that the field overall does not generate above average public interest. No second-level MeSH terms are cancer-specific, with the closest being *Neoplasms* (not significant), which is general enough to code both general and specialist interest papers.

From Model 2 (preliminary to addressing RQ2 in the next paragraph), despite the few positive results for RQ1, many more fields and many MeSH terms associated with different expected numbers of logged citations after considering logged Facebook mentions in Model 2. These results can be interpreted as field and MeSH term differences in citation rates. The significant MeSH terms could point to high citation specialties within a Science-Metrix field, substantial numbers of papers in general or multidisciplinary journals from a field with a different citation rate, or cross-field high/low citation topics. For example, the *Humanities* MeSH term associating with fewer logged citations might be due to humanities-oriented research in non-humanities journals being less cited (Cantín, Muñoz, & Roa, 2015). Most of the MeSH terms could describe academic specialties, with the possible exceptions of *Body Regions*. For example, *Immunological Phenomena* and *Immune System Phenomena* MeSH presumably relate to the **Immunology** Science-Metrix field (n=4771 papers in the dataset), and may identify immunology papers in general or multidisciplinary journals. Even the *Body Regions* category could relate to the **Anatomy & Morphology** Science-Metrix field (n=549) so the MeSH results do not give robust evidence of cross-field high citation specialties. Thus, the non-interaction field and MeSH terms in Model 2 add nothing to the long-established fact of field differences in citation rates. Nevertheless, they suggest that MeSH terms may supplement Science-Metrix field classifications by identifying the fields of some articles in general or multidisciplinary journals (a side effect rather than a goal of this study).

RQ2a (fields influencing the relationship between Facebook mentions and citations): The interaction terms in Model 2 address RQ2. Facebook mentions positively but weakly associated with citations in Model 2, confirming previous studies using simple correlations (Costas et al., 2015; Thelwall et al., 2013). From the Model 2 field interaction terms, the strongest relationship between citations and Facebook mentions is for **General and Internal Medicine**, which has high visibility on Facebook (Haustein et al., 2014). **Cardiovascular System and Hematology** is another field with stronger correlations between Facebook mentions and citations, perhaps because

Facebook is a popular for cardiologists and for patients to connect and share information (Kuehn, 2019), increasing the amount of data in the model. Similarly, **Developmental Biology** generates interest amongst the public and politicians (Losos et al., 2013). Conversely, in **General Science and Technology** and **Biochemistry and Molecular Biology**, Facebook mentions had a weaker relationship with citations than average. The first field seems likely to attract both citations and Facebook interest, so this suggests a mismatch between citable and public interest or publicity-generating papers in this field, perhaps because it is multidisciplinary. The latter case may be due to a lack of public interest or publicity for this field, compared to more directly health-related or general science fields.

RQ2b (MeSH terms influencing the relationship between Facebook mentions and citations): From the Model 2 MeSH interaction terms, most MeSH terms for which Facebook associates more strongly with citations are related to chemicals and drugs (e.g., *Biomedical and Dental Materials, Chemical Actions and Uses*) and biological sciences (e.g., *Genetic Phenomena*). This may be due to high citation rates and high Facebook visibility (Xia et al., 2016) for these fields, so the data is less noisy than for fields with lower values. *Geographic Locations* is a cross-field term associated with a weaker relationship between Facebook mentions and citation counts. This suggests that Facebook mentions are less useful for predicting citation counts when any medical or health topic is specific to a geographic location. This is plausible because Facebook is banned in China and Iran so Facebook mentions of research specific to these countries would probably be rare irrespective of its likelihood to be cited. The substantial research output from China would therefore disrupt the Facebook-citations relationship for location-specific research. To illustrate this, in October 2019 the 2016 article, “Long-term trend and spatial pattern of PM2.5 induced premature mortality in China” had 54 citations (from articles with at least one Chinese co-author) but no Facebook mentions, presumably because its information was primarily of interest in China. Similarly, the weaker Facebook-citations relationship for research with a MeSH classification of *Persons* may be due to some groups of people attracting more public interest on Facebook than others (e.g., using third-level MeSH terms: Athletes, Famous Persons vs. Working Poor, Homebound Persons).

6. Conclusions

In answer to the first research question, this article found only four (Science-Metrix) field differences and no (MeSH) cross-field topic differences in the rate of attracting Facebook mentions, although this is affected by the relatively strict statistical model used that filtered out minor relationships (Model 1), and the exclusion from the dataset of papers not mentioned on Facebook. The scarcity of statistically significant differences is surprising, given that previous studies of press coverage of academic research have found clinical medical topics to be the most extensively covered and some issues, such as cancer, to be commonly discussed (Lewison, Tootell, Roe, & Sullivan, 2008). A possible explanation is that a low proportion of papers generate substantial public interest in news or on Facebook, even for popular topics, and the attention given to the few is not enough to be statistically significant. The log transformations used in the models do not allow individual highly mentioned articles to dominate the results, such as those with extensive press coverage. This suggests that public interest in medical and health research topics tends to be limited in depth or breadth rather than field-wide. For example, the public might be more interested in breast cancer (a fourth level MeSH term: C04.588.180) than other types, such as eye cancer (C04.588.364), or might be mainly interested in studies suggesting a cure ignoring

other papers. The public might also post instead to online venues for patients and families to exchange medical and health information, such as PatientsLikeMe. Thus, it is not clear whether the scarcity of significant field differences and absence of significant MeSH differences is due to broadly equal Facebook interests for medical or health studies. Nevertheless, this study provides evidence of substantial Facebook attention, presumably reflecting public interest, for the first time at the subject level: **General Science and Technology**, **General and Internal Medicine**, **Complementary and Alternative Medicine**, and **Sport Sciences**.

The citation modelling considering Facebook mentions (Model 2) confirms, with a new dataset and a new modelling perspective, that Facebook mentions associate positively but weakly with citation counts which is in agreement with previous studies (Haustein et al., 2015; Thelwall et al., 2013).

For the second research question, five fields from Model 2 (including two of the same fields from Model 1) have statistically significantly different relationships between citations and Facebook mentions. One of the fields with many Facebook mentions also has a stronger positive relationship between mentions and citations (**General and Internal Medicine**), so papers in this field frequently mentioned on Facebook are more likely to be highly cited than average. In contrast, although papers in **General Science and Technology**, attract many Facebook mentions, these align more weakly with citation counts than average. This suggests a strong interest mismatch between the public (assuming that academics are not the dominant Facebook posters) and academics for this field (e.g., new technologies vs. methodological papers) that would also undermine any tendency for both constituencies to be more interested in higher quality or more useful studies. The MeSH terms provided the strongest evidence of mismatches between topics of public interest and topics of academic interest in terms of *Geographic Locations* and *Persons*. This means that Facebook mentions are less useful to predict future citation counts for topics with a variable level of public interest, and the Facebook user constituency should be considered when interpreting Facebook mention counts as indicators of public interest.

This research has some limitations. First, the data sources used for Facebook mentions, altmetric.com, is limited to the public Facebook posts while a recent study showed more than half of PLOS ONE papers were shared in non-public Facebook pages (Enkhbayar et al., 2019). Although the sample in this study is large, it does not cover all biomedical papers shared on Facebook. Moreover, since the sample is limited to biomedical and health sciences papers, the results do not apply to other fields. Additionally, different services to capture social media mentions of academic papers do not give the same data (Ortega, 2018), so the numbers would have been different if obtained from Plum Analytics, for example. Finally, some Facebook posts might be criticizing a paper, complicating the relationship between sharing and citation counts.

7. Theoretical and Practical Implications

Previous studies of the relationship between Facebook mentions and citation counts have been limited to simple correlations (Thelwall et al., 2013; Costas et al., 2015) or limited to samples of papers published in Nature (Xia et al., 2016) or Psychology (Vogl et al., 2018). This study includes a large and diverse sample of biomedical papers and uses novel statistical models to shed light on the role of Facebook in scholarly communication. The results suggest that attempts to use Facebook mentions as an indicator of public interest in research fields or topics should consider the Facebook constituency (e.g., few Chinese users), and the possibility that the public might not

be interested in all types of article in a field (e.g., cures vs. methods) or all topics within a field (e.g., breast cancer vs. eye cancer). These same considerations should apply when attempting to use early Facebook mentions as an indicator of longer-term citations. Moreover, this study revealed that medical and health related topics of scientific information shared on Facebook are different from traditional media and it may be beneficial for medical researchers and practitioners to use Facebook in developing strategies for communicating their research to a wider audiences. This is particularly important because diverse users within and outside academia use Facebook to disseminate their research articles (Mohammadi et al., 2019).

A methodological implication is that MeSH terms can be used to capture areas of public interest that do not correspond to academic fields. Although no statistically significant results were found for MeSH for RQ1, the positive result for RQ2 gives evidence of the value of this approach. Enforcing the subject classifications to be applied before MeSH, as above, was useful to ensure that the MeSH results were genuinely new findings that could not have been obtained before.

References

- Abramo, G. (2018). Revisiting the scientometric conceptualization of impact and its measurement. *Journal of Informetrics*, 12(3), 590–597. <https://doi.org/10.1016/j.joi.2018.05.001>
- Alperin, J. P. (2015). Geographic variation in social media metrics: An analysis of Latin American journal articles. *Aslib Journal of Information Management*, 67(3), 289–304. <https://doi.org/10.1108/AJIM-12-2014-0176>
- Archambault, E., Beauchesne, O., & Caruso, J. (2011). Towards a Multilingual, Comprehensive and Open Scientific Journal Ontology. In J. Noyons, E; Ngulube, P; Leta (Ed.), *Proceedings of ISSI 2011: The 13th Conference of The International Society for Scientometrics and Informetrics, Vols 1 and 2* (pp. 66–77). Durban, South Africa.
- Barata, G., Shores, K., & Alperin, J. P. (2018). Local chatter or international buzz? Language differences on posts about Zika research on Twitter and Facebook. *PLoS ONE*, 13(1), 1–15. <https://doi.org/10.1371/journal.pone.0190482>
- Bartlett, C. (2002). What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers. *BMJ*, 325(7355), 81–84. <https://doi.org/10.1136/bmj.325.7355.81>
- Bornmann, L. (2013). What is societal impact of research and how can it be assessed? a literature survey. *Journal of the American Society for Information Science and Technology*, 64(2), 217–233. <https://doi.org/10.1002/asi.22803>
- Bowman, T. D. (2015). *Investigating the use of affordances and framing techniques by scholars to manage personal and professional impressions on Twitter*. Indiana University, Bloomington, IN.USA.
- Bühlmann, P., & Meier, L. (2008). Discussion of “One-step sparse estimates in nonconcave penalized likelihood models” (auths H. Zou and R. Li). *Annals of Statistics*, 36, 1534–1541. <https://doi.org/10.1002/9781118445112.stat08190>

- Cantín, M., Muñoz, M., & Roa, I. (2015). Comparison between Impact Factor, Eigenfactor Score, and SCImago Journal Rank Indicator in Anatomy and Morphology Journals. *International Journal of Morphology*, 33(3), 1183–1188. <https://doi.org/10.4067/s0717-95022015000300060>
- Chapman, D. (2009). Advanced search features of PubMed. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Académie Canadienne de Psychiatrie de l'enfant et de l'adolescent*, 18(1), 58–59. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19270851>
- Collins, P. A., Abelson, J., Pyman, H., & Lavis, J. N. (2006). Are we expecting too much from print media? An analysis of newspaper coverage of the 2002 Canadian healthcare reform debate. *Social Science & Medicine*, 63(1), 89–102. <https://doi.org/10.1016/j.socscimed.2005.12.012>
- Costas, R., Zahedi, Z., & Wouters, P. (2015). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003–2019. <https://doi.org/10.1002/asi.23309>
- Dehdarirad, T., Didegah, F., & Sotudeh, H. (2018). Which Type of Research is Cited More Often in Wikipedia? A Case Study of PubMed Research. In R. Costas, T. Franssen, & A. Yegros-Yegros (Eds.), *23rd International Conference on Science and Technology Indicators (STI 2018)*. Leiden, The Netherlands: Centre for Science and Technology Studies (CWTS).
- Didegah, F. (2014). *Factors associating with the future citation impact of published articles: a statistical modelling approach. PQDT - UK & Ireland*. University of Wolverhampton (United Kingdom), Ann Arbor. Retrieved from <https://login.pallas2.tcl.sc.edu/login?url=https://search.proquest.com/docview/1687708176?accountid=13965>
- Didegah, F., Bowman, T. D., & Holmberg, K. (2016). Increasing our Understanding of Altmetrics : Identifying Factors That Are Driving Both Citation and Altmetric Counts. In *iConference* (pp. 1–8).
- Didegah, F., Bowman, T. D., & Holmberg, K. (2018). On the differences between citations and altmetrics: An investigation of factors driving altmetrics versus citations for finnish articles. *Journal of the Association for Information Science and Technology*, 69(6), 832–843. <https://doi.org/10.1002/asi.23934>
- Enkhbayar, A., Haustein, S., Barata, G., & Alperin, J. P. (2019). How much research shared on Facebook is hidden from public view? A comparison of public and private online activity around PLOS ONE papers. Retrieved from <http://arxiv.org/abs/1909.01476>
- Eysenbach, G. (2011). Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. *Journal of Medical Internet Research*, 13(4), e123. <https://doi.org/10.2196/jmir.2012>
- Facebook. (2019). Facebook Stats. Retrieved from <https://newsroom.fb.com/company-info/>
- Fenner, M. (2013). What Can Article-Level Metrics Do for You? *PLoS Biology*, 11(10), e1001687. Retrieved from <http://dx.plos.org/10.1371/journal.pbio.1001687>
- Gómez, I., Bordons, M., Fernández, M. T., & Méndez, A. (1996). Coping with the problem of subject classification diversity. *Scientometrics*, 35(2), 223–235. <https://doi.org/10.1007/BF02018480>
- Haustein, S., Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, 65(4), 656–669. <https://doi.org/10.1002/asi.23101>

- Haustein, Stefanie, Costas, R., & Larivière, V. (2015). Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. *PLOS ONE*, 10(3), e0120495. <https://doi.org/10.1371/journal.pone.0120495>
- Haustein, Stefanie, Peters, I., Sugimoto, C. R., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology*, 65(4), 656–669. <https://doi.org/10.1002/asi.23101>
- Holmberg, K., Bowman, S., Bowman, T., Didegah, F., & Kortelainen, T. (2019). What Is Societal Impact and Where Do Altimetrics Fit into the Equation? *Journal of Altimetrics*, 2(1), 1–11. <https://doi.org/10.29024/joa.21>
- Internet World Stats. (2019). World Internet Users and 2019 Population Stats. Retrieved from <https://internetworldstats.com/stats.htm>
- Kamel Boulos, M. N., & Anderson, P. F. (2012). Preliminary survey of leading general medicine journals' use of Facebook and Twitter 1. *Journal of the Canadian Health Libraries Association*, 33(02), 38–47. <https://doi.org/10.5596/c2012-010>
- Kastrati, Z., Imran, A. S., & Yayilgan, S. Y. (2019). The impact of deep learning on document classification using semantically rich representations. *Information Processing & Management*, 56(5), 1618–1632. <https://doi.org/10.1016/j.ipm.2019.05.003>
- Kousha, K., & Thelwall, M. (2019). An Automatic Method to Identify Citations to Journals in News Stories: A Case Study of UK Newspapers Citing Web of Science Journals. *Journal of Data and Information Science*, 4(3), 73–95. <https://doi.org/10.2478/jdis-2019-0016>
- Kuehn, B. M. (2019). Social Media Becomes a Growing Force in Cardiology. *Circulation*, 140(9), 790–792. <https://doi.org/10.1161/CIRCULATIONAHA.119.042813>
- Leimu, R., & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution*, 20(1), 28–32. <https://doi.org/10.1016/j.tree.2004.10.010>
- Lewison, G., Roe, P., Wentworth, A., & Szmukler, G. (2012). The reporting of mental disorders research in British media. *Psychological Medicine*, 42(2), 435–441. <https://doi.org/10.1017/S0033291711001012>
- Lewison, G., Tootell, S., Roe, P., & Sullivan, R. (2008). How do the media report cancer research? A study of the UK's BBC website. *British Journal of Cancer*, 99(4), 569–576. <https://doi.org/10.1038/sj.bjc.6604531>
- Lewison, Grant. (2002). From biomedical research to health improvement. *Scientometrics*, 54(2), 179–192. <https://doi.org/10.1023/A:1016005710371>
- Lewison, Grant, & Sullivan, R. (2008). The impact of cancer research: how publications influence UK cancer clinical guidelines. *British Journal of Cancer*, 98(12), 1944–1950. <https://doi.org/10.1038/sj.bjc.6604405>
- Lewison, Grant, & Turnbull, T. (2010). News in brief and features in New Scientist magazine and the biomedical research papers that they cite, August 2008 to July 2009. *Scientometrics*, 85(1), 345–359. <https://doi.org/10.1007/s11192-010-0221-9>
- Leydesdorff, L., Comins, J. A., Sorensen, A. A., Bornmann, L., & Hellsten, I. (2016). Cited references and Medical Subject Headings (MeSH) as two different knowledge representations: clustering and mappings at the paper level. *Scientometrics*, 109(3), 2077–2091. <https://doi.org/10.1007/s11192->

- Leydesdorff, L., & Opthof, T. (2013). Citation analysis with medical subject Headings (MeSH) using the Web of Knowledge: A new routine. *Journal of the American Society for Information Science and Technology*, 64(5), 1076–1080. <https://doi.org/10.1002/asi.22770>
- Liu, Y., & Wacholder, N. (2017). Evaluating the impact of MeSH (Medical Subject Headings) terms on different types of searchers. *Information Processing and Management*, 53(4), 851–870. <https://doi.org/10.1016/j.ipm.2017.03.004>
- Losos, J. B., Arnold, S. J., Bejerano, G., Iii, E. D. B., Hibbett, D., Moritz, C., ... Mindell, D. P. (2013). Evolutionary Biology for the 21st Century. *PLoS Biology*, 11(1). <https://doi.org/10.1371/journal.pbio.1001466>
- Mohammadi, E., Barahmand, N., & Thelwall, M. (2019). Who shares health and medical scholarly articles on Facebook? *Learned Publishing*, (November), 1–8. <https://doi.org/10.1002/leap.1271>
- Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K. L. (2018). Academic information on Twitter: A user survey. *PLOS ONE*, 13(5), e0197265. <https://doi.org/10.1371/journal.pone.0197265>
- Na, J. C., & Ye, Y. E. (2017). Content analysis of scholarly discussions of psychological academic articles on Facebook. *Online Information Review*, 41(3), 337–353. <https://doi.org/10.1108/OIR-02-2016-0058>
- National Institutes of Health. (2019). Impact of NIH Research. Retrieved from <https://www.nih.gov/about-nih/what-we-do/impact-nih-research>
- Noyons, E. (2019). Measuring Societal Impact Is as Complex as ABC. *Journal of Data and Information Science*, 4(3), 6–21. <https://doi.org/10.2478/jdis-2019-0012>
- Ortega, J. L. (2018). Reliability and accuracy of altmetric providers: a comparison among Altmetric.com, PlumX and Crossref Event Data. *Scientometrics*, 116, 2123–2138. <https://doi.org/10.1007/s11192-018-2838-z>
- Pallari, E., Lewison, G., Ciani, O., Tarricone, R., Sommariva, S., Begum, M., & Sullivan, R. (2018). The impacts of diabetes research from 31 European Countries in 2002 to 2013. *Research Evaluation*, 27(July), 270–282. <https://doi.org/10.1093/reseval/rvy006>
- Persson, O. (2010). Are highly cited papers more international? *Scientometrics*, 83(2), 397–401. <https://doi.org/10.1007/s11192-009-0007-0>
- Rees, C. E., & Bath, P. A. (2000). Mass media sources for breast cancer information: their advantages and disadvantages for women with the disease. *Journal of Documentation*, 56(3), 235–249. <https://doi.org/10.1108/EUM0000000007114>
- Ringelhan, S., Wollersheim, J., & Welp, I. M. (2015). I like, I cite? Do Facebook likes predict the impact of scientific work? *PLoS ONE*, 10(8), 1–21. <https://doi.org/10.1371/journal.pone.0134389>
- Sharma, V., Holmes, J., & Sarkar, I. (2016). Identifying Complementary and Alternative Medicine Usage Information from Internet Resources. *Methods of Information in Medicine*, 55(04), 322–332. <https://doi.org/10.3414/ME15-01-0154>
- Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology*, 65(5), 1018–1027. <https://doi.org/10.1002/asi.23037>

- Shu, F., Julien, C., Zhang, L., Qiu, J., Zhang, J., & Larivière, V. (2019). Comparing journal and paper level classifications of science. *Journal of Informetrics*, 13(1), 202–225. <https://doi.org/10.1016/j.joi.2018.12.005>
- Stremersch, S., Verniers, I., & Verhoef, P. C. (2007). The Quest for Citations: Drivers of Article Impact. *Journal of Marketing*, 71(3), 171–193. <https://doi.org/10.1509/jmkg.71.3.171>
- Sud, P., & Thelwall, M. (2014). Evaluating altmetrics. *Scientometrics*, 98(2), 1131–1143. <https://doi.org/10.1007/s11192-013-1117-2>
- Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of Informetrics*, 12(2), 430–435. <https://doi.org/10.1016/j.joi.2018.03.006>
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLoS ONE*, 8(5), e64841. <https://doi.org/10.1371/journal.pone.0064841>
- Thelwall, M., & Nevill, T. (2018). Could scientists use Altmeter.com scores to predict longer term citation counts? *Journal of Informetrics*, 12(1), 237–248. <https://doi.org/10.1016/j.joi.2018.01.008>
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Vogl, S., Scherndl, T., & Kühberger, A. (2018). #Psychology: a bibliometric analysis of psychological literature in the online media. *Scientometrics*, 115(3), 1253–1269. <https://doi.org/10.1007/s11192-018-2727-5>
- Weitkamp, E. (2003). British newspapers privilege health and medicine topics over other science news. *Public Relations Review*, 29(3), 321–333. [https://doi.org/10.1016/S0363-8111\(03\)00041-9](https://doi.org/10.1016/S0363-8111(03)00041-9)
- Williams, J. H. (2011). Use of social media to communicate sport science research. *International Journal of Sports Science and Coaching*, 6(2), 295–300. <https://doi.org/10.1260/1747-9541.6.2.295>
- Xia, F., Su, X., Wang, W., Zhang, C., Ning, Z., & Lee, I. (2016). Bibliographic analysis of Nature based on Twitter and Facebook altmetrics data. *PLoS ONE*, 11(12), 1–14. <https://doi.org/10.1371/journal.pone.0165997>
- Zahedi, Z., & Costas, R. (2018). General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PLoS ONE*, 13(5). <https://doi.org/10.1371/journal.pone.0197326>
- Zedda, M., & Barbaro, A. (2015). Adoption of Web 2.0 tools among STM publishers. How social are scientific journals? *Journal of the European Association for Health Information and Libraries*, 11, 9–12.
- Zheng, H., Aung, H. H., Erdt, M., Peng, T., Lansing, E., Raamkumar, A. S., & Theng, Y. (2019). Social Media Presence of Scholarly Journals. *Journal of the Association for Information Science and Technology*, 70(3), 256–270. <https://doi.org/10.1002/asi.24124>

Appendix 1: Table 2. Number of records in the final data set by Science-Metrix subfield

Subfield	Articles
Neurology & Neurosurgery	8031
General & Internal Medicine	7209
Psychiatry	5638
Cardiovascular System & Hematology	5057
General Science & Technology	4690
Developmental Biology	4338
Oncology & Carcinogenesis	4153
Immunology	3820
Obstetrics & Reproductive Medicine	3489
Public Health	3469
Orthopedics	3194
Gastroenterology & Hepatology	3086
Urology & Nephrology	3085
Microbiology	3076
Endocrinology & Metabolism	2770
Pediatrics	2598
Surgery	2519
Sport Sciences	2507
Biochemistry & Molecular Biology	2464
Emergency & Critical Care Medicine	2364
Nutrition & Dietetics	2242
Nursing	2226
Plant Biology & Botany	2166
Nuclear Medicine & Medical Imaging	2093
Health Policy & Services	1896
Anesthesiology	1803
Organic Chemistry	1764
Pharmacology & Pharmacy	1733
Experimental Psychology	1624
Respiratory System	1562
Rehabilitation	1512
Dermatology & Venereal Diseases	1452
Developmental & Child Psychology	1421
Ophthalmology & Optometry	1372
Substance Abuse	1341
Dentistry	1326
Pathology	1304
Allergy	1242
Epidemiology	1151
Evolutionary Biology	1147

Arthritis & Rheumatology	1130
Ecology	1129
Geriatrics	1121
Physiology	1105
Clinical Psychology	968
Veterinary Sciences	951
Virology	932
Genetics & Heredity	899
Toxicology	838
Environmental Sciences	794
Medicinal & Biomolecular Chemistry	786
Analytical Chemistry	725
Otorhinolaryngology	673
Bioinformatics	644
Gerontology	611
Behavioral Science & Comparative Psychology	568
Complementary & Alternative Medicine	560
Biotechnology	550
Medical Informatics	550
Biomedical Engineering	535
Environmental & Occupational Health	532
Food Science	509
Nanoscience & Nanotechnology	478
Anatomy & Morphology	478
General Chemistry	460
Tropical Medicine	392
Speech-Language Pathology & Audiology	384
Fisheries	324
Dairy & Animal Science	304
Social Psychology	302
Biophysics	300
Applied Ethics	272
Mycology & Parasitology	270
Chemical Physics	251
General Clinical Medicine	248
Anthropology	205
Entomology	200
Legal & Forensic Medicine	187
Polymers	173
Marine Biology & Hydrobiology	163
Optoelectronics & Photonics	160

Criminology	133
Strategic, Defence & Security Studies	118
Environmental Engineering	83
Human Factors	79
Acoustics	73
Logistics & Transportation	67
Education	66
Statistics & Probability	59
Science Studies	56
Fluids & Plasmas	54
History of Science, Technology & Medicine	53
General Arts, Humanities & Social Sciences	51
Inorganic & Nuclear Chemistry	48
Zoology	42
Psychoanalysis	39
Demography	38
Family Studies	34
Astronomy & Astrophysics	34
Optics	33
Sociology	32
Meteorology & Atmospheric Sciences	30
Gender Studies	29
Artificial Intelligence & Image Processing	27
Information & Library Sciences	25
Microscopy	24
Building & Construction	22
Law	16
Applied Physics	16
General Physics	14
Social Sciences Methods	13
Operations Research	10
Geochemistry & Geophysics	7
Agronomy & Agriculture	7
General Psychology & Cognitive Sciences	7
Social Work	5
Software Engineering	3
Physical Chemistry	3
Economics	2
Information Systems	1
Geology	1
Electrical & Electronic Engineering	1

Forestry	1
Energy	1
Total	134, 052

Appendix 2: Table 3. Number of articles in the final data set by 2nd level MeSH term.

Level 2 Subject	Articles
Eukaryota [B01]	191422
Investigative Techniques [E05]	125075
Environment and Public Health [N06]	102485
Persons [M01]	95956
Health Care Quality, Access, and Evaluation [N05]	94740
Diagnosis [E01]	69847
Pathological Conditions, Signs and Symptoms [C23]	49156
Amino Acids, Peptides, and Proteins [D12]	49121
Geographic Locations [Z01]	46421
Physiological Phenomena [G07]	45372
Therapeutics [E02]	43997
Behavior and Behavior Mechanisms [F01]	41657
Health Services Administration [N04]	40063
Chemical Actions and Uses [D27]	35770
Genetic Phenomena [G05]	33149
Health Care Facilities, Manpower, and Services [N02]	30845
Information Science [L01]	30811
Cells [A11]	30200
Social Sciences [I01]	28996
Chemical Phenomena [G02]	28872
Population Characteristics [N01]	27486
Psychological Phenomena [F02]	24219
Physical Phenomena [G01]	22961
Biological Factors [D23]	22722
Organic Chemicals [D02]	22302
Biological Phenomena [G16]	21853
Surgical Procedures, Operative [E04]	21672
Musculoskeletal and Neural Physiological Phenomena [G11]	21066
Natural Science Disciplines [H01]	20397
Cell Physiological Phenomena [G04]	19341

Neoplasms [C04]	19176
Enzymes and Coenzymes [D08]	18680
Nervous System Diseases [C10]	18318
Metabolism [G03]	18115
Mental Disorders [F03]	17721
Technology, Industry, and Agriculture [J01]	16344
Cardiovascular Diseases [C14]	16069
Heterocyclic Compounds [D03]	15787
Health Care Economics and Organizations [N03]	15660
Nervous System [A08]	15288
Nutritional and Metabolic Diseases [C18]	14506
Inorganic Chemicals [D01]	13381
Reproductive and Urinary Physiological Phenomena [G08]	13346
Health Occupations [H02]	13288
Female Urogenital Diseases and Pregnancy Complications [C13]	11621
Immune System Diseases [C20]	11495
Tissues [A10]	11429
Nucleic Acids, Nucleotides, and Nucleosides [D13]	11201
Digestive System Diseases [C06]	10817
Human Activities [I03]	10761
Behavioral Disciplines and Activities [F04]	10575
Mathematical Concepts [G17]	10401
Musculoskeletal System [A02]	10324
Respiratory Tract Diseases [C08]	9993
Humanities [K01]	9589
Carbohydrates [D09]	9325
Bacterial Infections and Mycoses [C01]	9298
Skin and Connective Tissue Diseases [C17]	9087
Circulatory and Respiratory Physiological Phenomena [G09]	9045
Education [I02]	8738
Food and Beverages [J02]	8719
Equipment and Supplies [E07]	8557
Polycyclic Compounds [D04]	8519
Virus Diseases [C02]	8505
Musculoskeletal Diseases [C05]	8438
Hemic and Immune Systems [A15]	8179
Congenital, Hereditary, and Neonatal Diseases and Abnormalities [C16]	8124
Male Urogenital Diseases [C12]	7986
Complex Mixtures [D20]	7890
Endocrine System Diseases [C19]	7702
Hormones, Hormone Substitutes, and Hormone Antagonists [D06]	7236

Lipids [D10]	7023
Wounds and Injuries [C26]	6973
Animal Diseases [C22]	6781
Bacteria [B03]	6444
Macromolecular Substances [D05]	6193
Microbiological Phenomena [G06]	6105
Body Regions [A01]	6005
Digestive System [A03]	5842
Viruses [B04]	5469
Cardiovascular System [A07]	5281
Pharmaceutical Preparations [D26]	5232
Immune System Phenomena [G12]	4774
Hemic and Lymphatic Diseases [C15]	4632
Chemically-Induced Disorders [C25]	4566
Urogenital System [A05]	4279
Plant Structures [A18]	3494
Fluids and Secretions [A12]	3485
Eye Diseases [C11]	3383
Biomedical and Dental Materials [D25]	2890
Parasitic Diseases [C03]	2741
Non-Medical Public and Private Facilities [J03]	2679
Organism Forms [B05]	2601
Respiratory System [A04]	2487
Sense Organs [A09]	2346
Otorhinolaryngologic Diseases [C09]	2333
Embryonic Structures [A16]	2219
Stomatognathic Diseases [C07]	2170
Digestive System and Oral Physiological Phenomena [G10]	1810
Stomatognathic System [A14]	1796
Ocular Physiological Phenomena [G14]	1667
Endocrine System [A06]	1648
Integumentary System [A17]	1558
Plant Physiological Phenomena [G15]	1545
Animal Structures [A13]	1487
Anesthesia and Analgesia [E03]	1431
Dentistry [E06]	1369
Integumentary System Physiological Phenomena [G13]	523
Bacterial Structures [A20]	458
Occupational Diseases [C24]	369
Fungal Structures [A19]	283
Viral Structures [A21]	115

Archaea [B02]	85
Disorders of Environmental Origin [C21]	26

Appendix 3: Tables corresponding to Figures 1 to 4.

Table 4. The values depicted in Figure 1, with Bonferroni-adjusted p-values included.

	point estimate	lower	upper	Bonf. p- value
Sport Sciences	42.56	12.04	73.09	<0.0001
Complementary and Alternative Medicine	54.52	22.89	86.14	<0.0001
General and Internal Medicine	58.08	27.76	88.39	<0.0001
General Science and Technology	76.82	46.44	107.19	<0.0001

Table 5. The values depicted in Figure 2, with Bonferroni-adjusted p-values included.

	point estimate	lower 99%	upper 99%	Bonf. p- value
Humanities	-20.30	-26.51	-14.09	<0.0001
Body Regions	-14.08	-21.45	-6.71	<0.0001
Physiological Phenomena	5.94	2.49	9.39	<0.0001
Amino Acids, Peptides, and Proteins	6.17	-0.59	12.93	0.0493
Investigative Techniques	6.58	0.74	12.42	0.0008
Psychological Phenomena	7.50	3.23	11.76	<0.0001
Therapeutics	8.68	5.37	11.98	<0.0001
Metabolism	9.77	4.33	15.21	<0.0001
Inorganic Chemicals	9.82	4.38	15.25	<0.0001
Health Care Quality, Access, and Evaluation	9.82	5.30	14.34	<0.0001
Health Services Administration	9.90	6.15	13.64	<0.0001
Mental Disorders	12.19	6.86	17.51	<0.0001
Cells	13.87	9.11	18.63	<0.0001
Cell Physiological Phenomena	13.94	8.34	19.53	<0.0001
Chemical Phenomena	14.32	9.23	19.41	<0.0001
Environment and Public Health	16.34	12.04	20.65	<0.0001
Nervous System	19.23	13.71	24.75	<0.0001
Immune System Phenomena	26.04	17.42	34.66	<0.0001
Microbiological Phenomena	32.29	23.84	40.73	<0.0001

Table 6. T values depicted in Figure 3, with Bonferroni-adjusted p-values included.

	point estimate	lower 99%	upper 99%	Bonf. p- value
Biochemistry and Molecular Biology	-0.32	-0.51	-0.12	<0.0001
General Science and Technology	-0.28	-0.36	-0.20	<0.0001
Developmental Biology	0.21	0.08	0.34	<0.0001
Cardiovascular System and Hematology	0.22	0.09	0.35	<0.0001
General and Internal Medicine	0.38	0.31	0.45	<0.0001

Table 7. The values depicted in Figure 4, with Bonferroni-adjusted p-values included.

	point estimate	lower 99%	upper 99%	Bonf. p- value
Geographic Locations	-0.08	-0.11	-0.05	<0.0001
Persons	-0.07	-0.09	-0.04	<0.0001
Information Science	0.06	0.03	0.09	<0.0001
Biological Phenomena	0.07	0.03	0.11	<0.0001
Pathological Conditions, Signs and Symptoms	0.07	0.04	0.10	<0.0001
Amino Acids, Peptides, and Proteins	0.08	0.02	0.14	<0.0001
Investigative Techniques	0.09	0.04	0.13	<0.0001
Eukaryota	0.10	0.04	0.15	<0.0001
Chemical Actions and Uses	0.10	0.07	0.13	<0.0001
Nucleic Acids, Nucleotides, and Nucleosides	0.11	0.05	0.16	<0.0001
Population Characteristics	0.12	0.08	0.15	<0.0001
Genetic Phenomena	0.12	0.09	0.16	<0.0001
Biomedical and Dental Materials	0.18	0.07	0.30	<0.0001